

An Automatic Method for Retrieving and Indexing Catalogues of Biomedical Courses

Victor Maojo, MD, PhD, Guillermo de la Calle, Miguel García-Remesal, PhD, Vaida Bankauskaite, MD, PhD, Jose Crespo, PhD

Biomedical Informatics Group, Universidad Politécnica de Madrid, Madrid, Spain

Abstract

Although there is wide information about Biomedical Informatics education and courses in different Websites, information is usually not exhaustive and difficult to update. We propose a new methodology based on information retrieval techniques for extracting, indexing and retrieving automatically information about educational offers. A web application has been developed to make available such information in an inventory of courses and educational offers.

Introduction

The authors are currently involved in a European-based initiative to disseminate and exchange results carried out in Biomedical Informatics (BMI) and Grid research and professional practice between the European Union, Latin America, the Western Balkans and North Africa. One of the goals of the project is to facilitate BMI and Grid training courses and opportunities for education exchange in the involved countries. To achieve this, we have developed a methodology to create automatically a biomedical informatics course repository.

Methods

Our methodology consists of several phases: first, a list of URLs containing information about courses is created. Then, Web pages are retrieved. Relevant data is extracted by using web/text mining and information retrieval techniques. A domain model is semiautomatically generated using this approach¹. Finally, course information is indexed and stored into a database. An application based on Web services technology has been developed to facilitate automatic searches. An earlier version of such information mediator has been published elsewhere².

Results

We have carried out an experiment using several websites to validate our approach. These sites contained information about courses in Europe and South America. A preliminary prototype of conceptual model for the BMI and Grid course

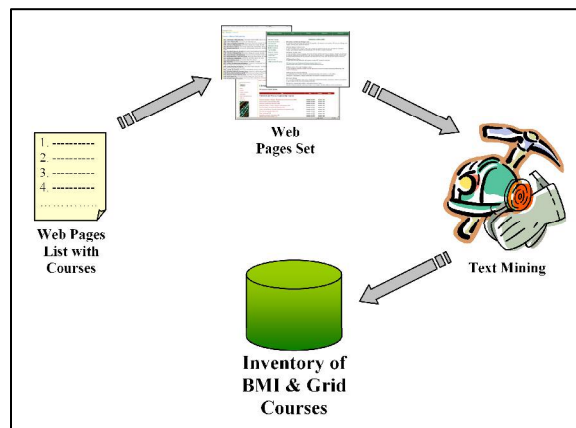


Figure 1. Methodology for creating course catalogue

domains is generated using this information. This schema consists of concepts and links between concepts for indexing contents and improving searches. A web application has been developed to access database's contents.

Conclusions and Future Work

By using the generated model, efficiency and accuracy in retrieving the information were improved. Current research plans to expand experiments adding more websites and extend it to catalogues of complete curricula in these fields.

Acknowledgement

This research has been funded by the Ministry of Education, Spain (TSI2006-13021-C02-01 and TIN2007-61768).

References

1. García-Remesal M, Maojo V, Crespo J et al. Logical Schema Acquisition from Text-Based Sources for Structured and Non-Structured Biomedical Sources Integration. AMIA Annu Symp Proc. 2007;259-263.
2. Van Mulligen EM, de la Calle G, Maojo V et al. Training Multidisciplinary Biomedical Informatics Students: Three Years of Experience. J Am Med Inform Assoc. 2008;15:246-254.